

EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	59	"edit distances" and (measure or measurement near3 similar or similarity)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:11
L2	31	"edit distances" and (measure or measurement near3 similar or similarity) and weight\$1	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:12
L3	0	"edit distances" and (measure or measurement near3 similar or similarity) and weight\$1 and (extract\$3 near3 vector\$1)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:13
L4	0	2 and (extract\$3 near3 vector)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:13
L5	304	"edit distance" and (measure or measurement near3 similar or similarity)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:14
L6	172	L5 and weight\$1	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:14
L7	4	L5 and weight\$1 and (extract\$3 near3 vector)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:14
L8	0	L5 and "similarity or similarities scores"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:15
L9	22	L5 and "similarity scores"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:16

EAST Search History

L10	10	L5 and (generat\$3 same vector\$1 same value!)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:16
L11	507	"707"/\$.ccls. and (generat\$3 same vector\$1 same value!)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:17
L12	1	"707"/\$.ccls. and ((generat\$3 same vector\$1 same value!) and ("similarity score" and "similarity vector"))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:18
L13	0	706/6.15,37,44,.ccls. and ((generat\$3 same vector\$1 same value!) and ("similarity score" and "similarity vector"))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:19
L14	0	706/6.15,37,44,.ccls. and ((generat\$3 same vector\$1 same value!) and ("similarity score" and "similarity vector") and (neural same network))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:20
L15	57	706/6.15,37,44,.ccls. and (neural same network)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:20
L16	16	706/6.15,37,44,.ccls. and (neural same network) and (vector\$1 same value!)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:21
L17	22	L5 and "similarity scores"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:21
L18	4	L17 and indicator	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:21
L19	4	"707"/\$.ccls. and ((generat\$3 same vector same value!) and ("edit distance" and (measure or measurement near3 similar or similarity)) and weight)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:22

EAST Search History

L20	1	("similarity score" and "similarity vector") and "measuring similarity"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:22
L21	1	"similarity score" and "similarity vector" and (generat\$3 same indicator)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:23
L22	19	"707"/\$.ccls. and (((("edit distance" or (edit same distance)) and (measure or measurement near3 similar or similarity)) and (generat\$3 or creat\$3 same vector same value!) and (extract\$3 same vector))	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:24
L23	22	((("edit distance" or (edit same distance)) and (measure or measurement near3 similar or similarity)) and "similarity scores"	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:24
L24	2	"706"/\$.ccls. and ((generat\$3 same vector same value!) and ((("edit distance" or (edit same distance)) and (measure or measurement near3 similar or similarity)) and weight)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:25
L25	4	"707"/\$.ccls. and ((generat\$3 same vector same value!) and ((("edit distance" or (edit same distance)) and (measure or measurement near3 similar or similarity)) and weight)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:25
L26	522	("edit distance" or (edit same distance)) and (measure or measurement near3 similar or similarity)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:25
L27	471	((("edit distance" or (edit same distance)) and (measure or measurement near3 similar or similarity)) and (generat\$3 or creat\$3 same vector same value!)	US-PGPUB; USPAT; EPO; JPO; DERWENT; IBM_TDB	OR	OFF	2007/07/05 12:26



USPTO

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

Search: ☒ The ACM Digital Library ☐ The Guide

training data set by matching pairs of data fields between each



[Feedback](#) [Report a problem](#) [Satisfaction survey](#)

Terms used:

training data set by matching pairs of data fields between each pairs of records

Found 151,839 of 205,978

Sort results
by

relevance

Display
results

expanded form



[Save results to a Binder](#)



[Search Tips](#)

☐ Open results in a new window

Try an [Advanced Search](#)

Try this search in [The ACM Guide](#)

Results 1 - 20 of 200

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

Best 200 shown

Relevance scale ☐ ☐ ☐ ☐ ☐

1 [Research track: Adaptive duplicate detection using learnable string similarity](#)



[measures](#)

Mikhail Bilenko, Raymond J. Mooney

August 2003 **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '03**

Publisher: ACM Press

Full text available: pdf(239.92 KB)

Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

The problem of identifying approximately duplicate records in databases is an essential step for data cleaning and data integration processes. Most existing approaches have relied on generic or manually tuned distance metrics for estimating the similarity of potential duplicates. In this paper, we present a framework for improving duplicate detection using trainable measures of textual similarity. We propose to employ learnable text distance functions for each database field, and show that such ...

Keywords: SVM applications, data cleaning, distance metric learning, record linkage, string edit distance, trained similarity measures

2 [Fast detection of communication patterns in distributed executions](#)



Thomas Kunz, Michiel F. H. Seuren

November 1997 **Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research CASCON '97**

Publisher: IBM Press

Full text available: pdf(4.21 MB)

Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Understanding distributed applications is a tedious and difficult task. Visualizations based on process-time diagrams are often used to obtain a better understanding of the execution of the application. The visualization tool we use is Poet, an event tracer developed at the University of Waterloo. However, these diagrams are often very complex and do not provide the user with the desired overview of the application. In our experience, such tools display repeated occurrences of non-trivial commun ...


3 [The relational model for database management: version 2](#)



E. F. Codd

January 1990 Book

Publisher: Addison-Wesley Longman Publishing Co., Inc.

Full text available:  pdf(28.61 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

From the Preface (See Front Matter for full Preface)

An important adjunct to precision is a sound theoretical foundation. The relational model is solidly based on two parts of mathematics: firstorder predicate logic and the theory of relations. This book, however, does not dwell on the theoretical foundations, but rather on all the features of the relational model that I now perceive as important for database users, and therefore for DBMS vendors. My perceptions result from 20 y ...

4 [Selected writings on computing: a personal perspective](#) 

Edsger W. Dijkstra
January 1982 Book

Publisher: Springer-Verlag New York, Inc.

Full text available:  pdf(60.98 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)

Since the summer of 1973, when I became a Burroughs Research Fellow, my life has been very different from what it had been before. The daily routine changed: instead of going to the University each day, where I used to spend most of my time in the company of others, I now went there only one day a week and was most of the time that is, when not travelling!-- alone in my study. In my solitude, mail and the written word in general became more and more important. The circumstance that my employe ...



5 [Exploiting perception in high-fidelity virtual environments: Exploiting perception in high-fidelity virtual environments](#) 



Additional presentations from the 24th course are available on the citation page

Mashhuda Glencross, Alan G. Chalmers, Ming C. Lin, Miguel A. Otaduy, Diego Gutierrez
July 2006 **ACM SIGGRAPH 2006 Courses SIGGRAPH '06**

Publisher: ACM Press

Full text available:  pdf(5.07 MB)  mov(68:6 MIN) Additional Information: [full citation](#), [appendices and supplements](#), [abstract](#), [references](#), [cited by](#), [index terms](#)


The objective of this course is to provide an introduction to the issues that must be considered when building high-fidelity 3D engaging shared virtual environments. The principles of human perception guide important development of algorithms and techniques in collaboration, graphical, auditory, and haptic rendering. We aim to show how human perception is exploited to achieve realism in high fidelity environments within the constraints of available finite computational resources. In this course w ...

Keywords: collaborative environments, haptics, high-fidelity rendering, human-computer interaction, multi-user, networked applications, perception, virtual reality

6 [Classics in software engineering](#) 

January 1979 Divisible Book

Publisher: Yourdon Press

Full text available:  pdf(22.45 MB) Additional Information: [full citation](#), [cited by](#), [index terms](#)

7 [Poster papers: Learning to match and cluster large high-dimensional data sets for data integration](#) 



William W. Cohen, Jacob Richman

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '02**

Publisher: ACM Press

Full text available: pdf(634.07 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Part of the process of data integration is determining which sets of identifiers refer to the same real-world entities. In integrating databases found on the Web or obtained by using information extraction methods, it is often possible to solve this problem by exploiting similarities in the textual names used for objects in different databases. In this paper we describe techniques for clustering and matching identifier names that are both scalable and *adaptive*, in the sense that they can ...

Keywords: clustering, large datasets, learning, text mining

8 Artificial intelligence

Elaine Rich

January 1983 Book

Publisher: McGraw-Hill, Inc.

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [review](#)

The goal of this book is to provide programmers and computer scientists with a readable introduction to the problems and techniques of artificial intelligence (A.I.). The book can be used either as a text for a course on A.I. or as a self-study guide for computer professionals who want to learn what A.I. is all about.

The book was designed as the text for a one-semester, introductory graduate course in A.I. In such a course, it should be possible to cover all of the material in the boo ...

9 Collective entity resolution in relational data



Indrajit Bhattacharya, Lise Getoor

March 2007 **ACM Transactions on Knowledge Discovery from Data (TKDD)**, Volume 1 Issue 1

Publisher: ACM Press

Full text available: pdf(511.57 KB) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

Many databases contain uncertain and imprecise references to real-world entities. The absence of identifiers for the underlying entities often results in a database which contains multiple references to the same entity. This can lead not only to data redundancy, but also inaccuracies in query processing and knowledge extraction. These problems can be alleviated through the use of *entity resolution*. Entity resolution involves discovering the underlying entities and mapping each database ...

Keywords: Entity resolution, data cleaning, graph clustering, record linkage

10 Learning methods: Interactive deduplication using active learning



Sunita Sarawagi, Anuradha Bhamidipaty

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining KDD '02**

Publisher: ACM Press

Full text available: pdf(1.14 MB) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Deduplication is a key operation in integrating data from multiple sources. The main challenge in this task is designing a function that can resolve when a pair of records refer

to the same entity in spite of various data inconsistencies. Most existing systems use hand-coded functions. One way to overcome the tedium of hand-coding is to train a classifier to distinguish between duplicates and non-duplicates. The success of this method critically hinges on being able to provide a *covering and ...*

11 A Bayesian decision model for cost optimal record matching

V. S. Verykios, G. V. Moustakides, M. G. Elfeky

May 2003 **The VLDB Journal – The International Journal on Very Large Data Bases**,
Volume 12 Issue 1

Publisher: Springer-Verlag New York, Inc.

Full text available:  [pdf\(180.87 KB\)](#) Additional Information: [full citation](#), [abstract](#), [citations](#), [index terms](#)

In an error-free system with perfectly clean data, the construction of a global view of the data consists of linking - in relational terms, joining - two or more tables on their key fields. Unfortunately, most of the time, these data are neither carefully controlled for quality nor necessarily defined commonly across different data sources. As a result, the creation of such a global data view resorts to approximate joins. In this paper, an optimal solution is proposed for the matching or the lin ...

Keywords: Cost optimal statistical model, Data cleaning, Record linkage

12 Anatomy of LISP

John Allen


January 1978 Book

Publisher: McGraw-Hill, Inc.

Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [index terms](#)


This text is nominally about LISP and data structures. However, in the process it covers much broader areas of computer science. The author has long felt that the beginning student of computer science has been getting 'a distorted and disjointed picture of the field. In some ways this confusion is natural; the field has been growing at such a rapid rate that few are prepared to be judged experts in all areas of the discipline. The current alternative seems to be to give a few introductory cou ...

13 Dynamic speculation and synchronization of data dependences

 Andreas Moshovos, Scott E. Breach, T. N. Vijaykumar, Gurindar S. Sohi

May 1997 **ACM SIGARCH Computer Architecture News , Proceedings of the 24th annual international symposium on Computer architecture ISCA '97**, Volume 25 Issue 2

Publisher: ACM Press

Full text available:  [pdf\(2.51 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#)

Data dependence speculation is used in instruction-level parallel (ILP) processors to allow early execution of an instruction before a logically preceding instruction on which it may be data dependent. If the instruction is independent, data dependence speculation succeeds; if not, it fails, and the two instructions must be synchronized. The modern dynamically scheduled processors that use data dependence speculation do so blindly (i.e., every load instruction with unresolved dependences is spec ...

14 Essays in computing science

C. A. R. Hoare

January 1989 Book

Publisher: Prentice-Hall, Inc.

Full text available:  [pdf\(20.91 MB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [cited by](#), [review](#)

Charles Antony Richard Hoare is one of the most productive and prolific computer scientists. This volume contains a selection of his published papers. There is a need, as in a Shakespearian Chorus, to offer some apology for what the book manifestly fails to achieve. It is not a complete 'collected works'. Selection between papers of this quality is not easy and, given the book's already considerable size, some difficult decisions as to what to omit have had to be made. Pity the editor weighin ...

15 Research sessions: Research 9: Schema matching: Multi-column substring matching for database schema translation

Robert H. Warren, Frank Wm. Tompa

September 2006 **Proceedings of the 32nd international conference on Very large data bases VLDB '06**

Publisher: VLDB Endowment

Full text available:  [pdf\(604.58 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [index terms](#)

We describe a method for discovering complex schema translations involving substrings from multiple database columns. The method does not require a training set of instances linked across databases and it is capable of dealing with both fixed-and variable-length field columns. We propose an iterative algorithm that deduces the correct sequence of concatenations of column substrings in order to translate from one database to another. We introduce the algorithm along with examples on common databa ...

16 Paper session II: record linkage, entity resolution: Blocking-aware private record linkage



Ali Al-Lawati, Dongwon Lee, Patrick McDaniel

June 2005 **Proceedings of the 2nd international workshop on Information quality in information systems IQIS '05**

Publisher: ACM Press

Full text available:  [pdf\(658.96 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#)

In this paper, the problem of quickly matching records (i.e., record linkage problem) from two autonomous sources without revealing privacy to the other parties is considered. In particular, our focus is to devise *secure blocking* scheme to improve the performance of record linkage significantly while being secure. Although there have been works on private record linkage, none has considered adopting the blocking framework. Therefore, our proposed blocking-aware private record linkage can ...

17 Data clustering: a review



A. K. Jain, M. N. Murty, P. J. Flynn

September 1999 **ACM Computing Surveys (CSUR)**, Volume 31 Issue 3

Publisher: ACM Press

Full text available:  [pdf\(636.24 KB\)](#) Additional Information: [full citation](#), [abstract](#), [references](#), [citations](#), [index terms](#), [review](#)

Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters). The clustering problem has been addressed in many contexts and by researchers in many disciplines; this reflects its broad appeal and usefulness as one of the steps in exploratory data analysis. However, clustering is a difficult problem combinatorially, and differences in assumptions and contexts in different communities has made the transfer of useful generic co ...

Keywords: cluster analysis, clustering applications, exploratory data analysis, incremental clustering, similarity indices, unsupervised learning

18

Projectors: advanced graphics and vision techniques



Ramesh Raskar

August 2004 **ACM SIGGRAPH 2004 Course Notes SIGGRAPH '04**

Publisher: ACM Press

Full text available: [pdf\(6.53 MB\)](#) Additional Information: [full citation](#)

19 Classification in Networked Data: A Toolkit and a Univariate Case Study

Sofus A. Macskassy, Foster Provost

May 2007 **The Journal of Machine Learning Research**, Volume 8

Publisher: MIT Press

Full text available: [pdf\(517.66 KB\)](#) Additional Information: [full citation](#), [abstract](#)

This paper is about classifying entities that are interlinked with entities for which the class is known. After surveying prior work, we present NetKit, a modular toolkit for classification in networked data, and a case-study of its application to networked data used in prior machine learning research. NetKit is based on a node-centric framework in which classifiers comprise a local classifier, a relational classifier, and a collective inference procedure. Various existing node-centric relational classifiers are compared.

20 Shape-based retrieval and analysis of 3D models



Thomas Funkhouser, Michael Kazhdan

August 2004 **ACM SIGGRAPH 2004 Course Notes SIGGRAPH '04**

Publisher: ACM Press

Full text available: [pdf\(12.56 MB\)](#) Additional Information: [full citation](#), [abstract](#)

Large repositories of 3D data are rapidly becoming available in several fields, including mechanical CAD, molecular biology, and computer graphics. As the number of 3D models grows, there is an increasing need for computer algorithms to help people find the interesting ones and discover relationships between them. Unfortunately, traditional text-based search techniques are not always effective for 3D models, especially when queries are geometric in nature (e.g., find me objects that fit into this space).

Results 1 - 20 of 200

Result page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

The ACM Portal is published by the Association for Computing Machinery. Copyright © 2007 ACM, Inc.

[Terms of Usage](#) [Privacy Policy](#) [Code of Ethics](#) [Contact Us](#)

Useful downloads: [Adobe Acrobat](#) [QuickTime](#) [Windows Media Player](#) [Real Player](#)